dbGaP Study Release Notes

Release Notes for NHLBI TOPMed - NHGRI CCDG CCDG BioVU, phs001624.v2.p2
"NHLBI TOPMed - NHGRI CCDG: Vanderbilt University BioVU Atrial Fibrillation Genetics Study"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

August	27, 2020	Version 1 Data set release date
September	09, 2021	Version 2 Data set release date

2021-09-09
Version 2 Data set release for NHLBI TOPMed - NHGRI CCDG BioVU now available

This release includes phenotype tables, the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (GSO) (HMB-GSO)

Data Type		Subjects	Samples
Phenotypes		2666	2739
Molecular Data	SNP/CNV Genotypes (NGS)*[1]	2656	2728
	WGS	2656	2728*[2]

*[1] For a description of SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

*[2] Additional 11 EA samples are not included.

Molecular Data (Genotype) Information

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
    a. phg001641.v1.TOPMed_CCDG_BioVU_v2_frz9.sample-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
    a. phg001641.v1.TOPMed_CCDG_BioVU_v2_frz9.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-GSO.tar.gz

Study and Phenotype Data Updates

- **New Study Accession**

NHLBI TOPMed - NHGRI CCDG BioVU version 1, phs001624.v1.p1, has been updated to version 2. The dbGaP accession for the current set of data is phs001624.v2.p2, and 915 new subjects and 988 new samples are added. The participant number (p#) has been changed in version 2, since 2 subjects and 2 samples have been retired from the previous version.

- **Updated datasets (n=4 datasets)**

pht	version	Dataset Name
8229	2	TOPMed_CCDG_BioVU_Subject
8230	2	TOPMed_CCDG_BioVU_Sample
8231	2	TOPMed_CCDG_BioVU_Subject_Phenotypes
8232	2	TOPMed_CCDG_BioVU_Sample_Attributes

- **New variables (n=1 variable)**

pht	pht version	Dataset Name	phv	Variable Name
8229	2	TOPMed_CCDG_BioVU_Subject	492548	SEX

- **Updated variables (n=25 variables)**

pht	pht version	phv	phv version	Variable Name
8229	2	389849	2	SUBJECT_ID
8229	2	389850	2	CONSENT
8229	2	389851	2	SUBJECT_SOURCE
8229	2	389852	2	SOURCE_SUBJECT_ID
8230	2	389854	2	SUBJECT_ID
8230	2	389855	2	SAMPLE_ID
8231	2	389859	2	SUBJECT_ID
8231	2	389860	2	GENDER
8231	2	431149	2	RACE
8231	2	431150	2	AGE
8231	2	431151	2	AGE_BASELINE
8231	2	431152	2	AGE_CVD
8231	2	431153	2	CVD_STATUS
8231	2	431154	2	AF_STATUS
8232	2	389861	2	SAMPLE_ID
8232	2	389862	2	BODY_SITE
8232	2	389863	2	ANALYTE_TYPE
8232	2	389864	2	SEQUENCING_CENTER
8232	2	431155	2	IS_TUMOR
8232	2	431156	2	Study_Name
8232	2	431157	2	TOPMed_Project
8232	2	431158	2	TOPMed_Phase
8232	2	431159	2	Funding_Source
8232	2	431160	2	CCDG_SAMPLE_ID
8232	2	431161	2	CCDG_SAMPLE_ID_alt

dbGaP Study Release Notes

Authorized Access (Individual Level Data and Sequence Data)

Individual level data and sequence data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and document will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001624/phs001624.v2.p1

2020-08-27

Version 1 Data set release for NHLBI TOPMed CCDG BioVU now available

This release includes updated phenotype tables, genotype files, and whole genome sequencing data brokered through the SRA. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (GSO) (HMB-GSO)

	Phenotype	Genotype	SRA*
subjects	1753	1129	1129
samples	1753	1129	1129

* Additional 624 EA samples are not included.

Molecular Data (Genotype) Information

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
    a. phg001451.v1.TOPMed_CCDG_BioVU.sample-info.MULTI.tar.gz
    b. phg001451.v1.TOPMed_CCDG_BioVU.marker-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
    a. phg001451.v1.TOPMed_CCDG_BioVU.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-GSO.tar.gz

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and document will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001624/phs001624.v1.p1